International Journal of Advanced Technology in Engineering and Science
Vol. No. 12, Issue No. 05, May 2024
www.ijates.com
ijates
ISSN 2348 - 7550

Predicting Stock Prices with SVM Algorithms by Using Machine Learning

Mrs. I. Padmaja[1], K. Nandini[2], Sk. Hussen Basha[3], Sk. Saifullah[4], L. Sandeep Kumar[5]

[1]Assistant Professor, Dept of Electronics and Communication Engineering, Tirumala Engineering College

[2,3,4,5] Ug Students, Dept of Electronics and Communication Engineering, Tirumala Engineering College.

[1]Padmajasdk@Gmail.Com, [2] Nandinikandula26@Gmail.Com, [3]Hussenbasha462@Gmail.Com,

[4] Saifullah04082001@Gmail.Com, [5]Sandeeplinga7273@Gmail.Com

Abstract:

Predicting Stock Prices through the lens of machine learning, specifically focusing onSupervised Machine (SVM) algorithms. Our research aims to investigate the efficiency of Support Vector Machine in comparison toRandom Forest algorithms for predicting stock prices. By harnessing the power of machine learning, we seek to explore which algorithmyields superior accuracy in forecasting stock market trends. SVM, known for its ability to handle high-dimensional data and nonlinear relationships, holds promise in capturing intricate patterns within stock data. Through rigorous experimentation and analysis, we aim to demonstrate that SVM outperforms than Random Forest algorithms in accurately predicting stock prices, thereby offering investors a more reliable tool for decision- making in financial markets. This research contributes valuable insights into the application of machine learning techniques in the domain ofstock market prediction, potentially paving the way for more robust and precise investment strategies.

Keywords: Data mining, Machine Learning, Stock Prediction, Data Cleaning, support Vector Machine,Random Forest algorithm, Data normalisation

I. INTRODUCTION

Predicting stock prices using supervised machine learning techniques like Support Vector Machines (SVM) and Random Forest is a complex but fascinating field. In this introduction, we'll break down the basics of how these methods work and how they can be applied to forecast stock prices.

Firstly, let's understand what supervised machine learning means. In supervised learning, we train a model using labelled data, which means the model learns from examples that have inputs and corresponding outputs. In the case of predicting stock prices, the inputs could be various factors likehistorical stock prices, trading volume, economic indicators, etc., while the output would be the

futurestock price.

Support Vector Machines (SVM) is a type of supervised learning algorithm used for classification and regression tasks. It works by finding the best possible line or hyperplane that separates different classes or, in the case of regression, predicts values based on input features. In stock price prediction, SVM can be trained to recognize patterns in historical data and make predictions based on those patterns.

Random Forest is another popular supervised learning algorithm used for both classification and regression tasks. It's an ensemble learning method that works by creating multiple decision trees during training and outputting the averageprediction of the individual trees. In the context of stock price prediction, Random Forest can be trained to consider various factors and make predictions based on the collective wisdom of the decision trees.

To use SVM or Random Forest for stock price prediction, we first need to gather relevant data. This includes historical stock prices, trading volume, company financials, economic indicators, and any other factors that might influence stock prices. Once we have the data, we preprocess it to remove noise and ensure it's in a format that the algorithms can understand.

Next, we split the data into training and testing sets.The training set is used to train the SVM or RandomForest model, while the testing set is used to evaluate its performance and see how well it generalizes to new, unseen data. It's important to usea separate testing set to avoid overfitting, where the model performs well on the training data but poorly on new data.

During the training phase, the SVM or Random Forest algorithm learns from the input features and their corresponding output values. It adjusts its internal parameters to minimize the error between the predicted values and the actual values in the training data. This process continues until the model converges to the best possible solution or reaches a stopping criterion.

Once the model is trained, we can use it to make predictions on new data. For stock price prediction, we input the relevant features (e.g., historical prices,trading volume) into the trained model, and it outputs a predicted stock price. Keep in mind that these predictions are not guaranteed to be accurate and should be used as guidance rather than definitive forecasts.

In conclusion, using supervised machine learning techniques like Support Vector Machines and Random Forest can help us predict stock prices by learning from historical data and identifying patterns that can be used to make future predictions.However, it's essential to gather high-quality

data, carefully preprocess it, and evaluate the model'sperformance to ensure reliable predictions.

II. LITERATURE SURVEY

Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend: Publication Year: 2018 Author: Jai Jawaani, Hardik Sachdeva, Manav Gupta, Alka Singhal Journal Name: 2018 IEEE Summary: To identify the

[2] relationship between different existing time series algorithms namely ARIMA and Holt Winter and the stock prices is the main objective of the proposed work, for the investments a good risk-free range of stock prices are analysed and therefore better accuracy of the model can be seen. To find distinguished results for shares in the stock market, the combination of two different time series analysismodels is opted by producing a range of prices to theconsumer of the stocks. Not complex in nature and estimation of values which are purely based on the past stock prices for non-seasonal or seasonal is the main advantage of these models. In this experiment,some limitations are, the work that never takes into consideration and other circumstances like news about any new market strategy or media release relevant to any company which may get affected by the prices of stocks.

Stock Price Prediction Based on Information Entropy and Artificial Neural Network: Publication Year: 2019 Author: Zang Yuze, Wang Yawing Journal Name: 2019 IEEE Summary: One of the most important components of the financial system is the stock market. [11] For supporting the activity and evolvement, money is directed by the investors of the associated firm. Along with information theoryand Artificial Neural Network (ANN) the combination of machine learning framework is formed. Information entropy for non-linear causalityand stock relevance also to facilitate ANN time series modelling are creatively used by this method. The feasibility of this machine learning framework is analysed with Amazon, Apple, Google and Facebookprices. A time series analysis method based on information theory as well as LSTM to model the stock price dynamics. The transfer entropy between relevant variables to help LSTM time series prediction is merged in this modelling infrastructure,the accuracy of the assumption outcome is broadly granted. Modelled and real stock price is highly correlated while differ slightly in terms of Mean Absolute Error.

III. METHODOLOGYCLASSIFIED ALGORITHMS

Classification algorithms are used when the target variable is categorical, meaning it falls into discrete classes or categories. Examples of classification algorithms include: Logistic Regression, Decision Trees, Support Vector Machines (SVM), Random Forest, K-Nearest Neighbours (KNN) Classification algorithms are commonly used for tasks such as spam detection, image recognition, and sentiment analysis. A classification algorithm is computational technique used in machine

learning to categorize data points into predefined classes or categories based on their features. One common algorithm is Decision Trees, which iteratively splits the data based on feature values to create a tree-like structure.

Logistic Regression

Logistic regression is a fundamental algorithm in machine learning used for binary classification tasks, where the goal is to predict the probability that an instance belongs to a certain class.

Despite its name, logistic regression is a classification algorithm rather than a regression one.The algorithm models the relationship between the independent variables and the binary outcome using the logistic function, also known as the sigmoidfunction. This function maps any real-valued number into a probability value between 0 and 1, making it suitable for binary classification.

One of the key advantages of logistic regression is its simplicity and interpretability. Unlike more complexalgorithms such as neural networks, logistic regression produces results that are easy to understand and interpret. Additionally, logistic regression performs well when the relationship between the independent variables and the log-odds of the dependent variable is approximately linear. However, logistic regression may not perform well ifthe relationship is highly nonlinear or if there are interactions between the independent variables that are not accounted for.Despite these limitations, logistic regression remains a powerful and widely used algorithm in the field of machine learning, particularly in cases where interpretability is important and the data exhibits a linear relationship between the independent variables and the log-odds of the dependent variable.

Decision Trees

Decision trees are versatile and intuitive machine learning algorithms used for both classification and regression tasks. They work by recursivelypartitioning the feature space into smaller regions based on the values of input features. At each step ofthe partitioning process, decision trees identify the feature that best splits the data into purest possible subsets, maximizing the homogeneity within each subset. This splitting criterion is typically based on metrics such as Gini impurity or entropy for classification tasks, and mean squared error for regression tasks.

The resulting tree structure resembles a flowchart, where each internal node represents a decision basedon a feature, each branch represents an outcome of that decision, and each leaf node represents the finalpredicted class or value. Decision trees are favouredfor their interpretability and ease of visualization, allowing users to understand the underlying decision-making process. However, decision trees are prone to overfitting, especially when they grow deep and complex. Techniques such as pruning, limiting the maximum depth of the tree, or setting a minimum number of samples required

to split a node can help mitigate overfitting and improve generalization performance. Additionally, ensemble methods like Random Forests and Gradient BoostingMachines are often employed to further enhancepredictive accuracy by combining multiple decision trees.

Support Vector Machine (SVM):

Imagine you have a bunch of dots on a piece of paper,and you want to draw a line to separate them into twogroups. That's what SVM does, but in many dimensions. It looks at historical data about stocks, like their past prices and trading volumes, and tries to draw the best line (or plane, if you're thinking in 3D) to separate the data into different categories. These categories might be things like "buy," "sell," or "hold." The cool thing about SVM is that it's reallygood at finding patterns in the data, even if those patterns aren't obvious to us humans. It's like havinga super-smart detective that can spot subtle clues in the stock market trends. By mapping historical stockdata onto a high-dimensional space and identifying the optimal hyperplane, SVMs effectively capture complex relationships between input features (such as historical price trends, trading volumes, and market indicators) and target variables (future stock prices)

Through this process of maximizing the margin between different classes while minimizing classification errors, SVMs can generalize well to unseen data, making them valuable tools for forecasting stock prices.

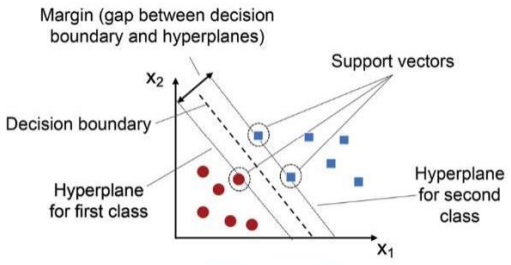

Fig (5): Support vector machine

Random Forest Classifier:

Now, imagine you're in a forest, and each tree represents a different way to predict stock prices. Random Forest is like asking a bunch of trees for their opinions and then combining them to make a decision. Each tree looks at different features of the stock data, like the company's earnings or industry trends, and makes its own prediction. Random Forestis great because it's like having a team of experts giving you advice. It’s robust and can handle lots of different scenarios without getting confused.

Random Forest Classifier is used for classification tasks in machine learning, leveraging an ensemble ofdecision trees to make predictions. It offers robustness against overfitting by aggregating multiple decision trees and can handle large datasets with high dimensionality. Random Forests excel in handling both numerical and categorical data, making them versatile and widely applicable in various domains such as finance, healthcare, andmarketing.

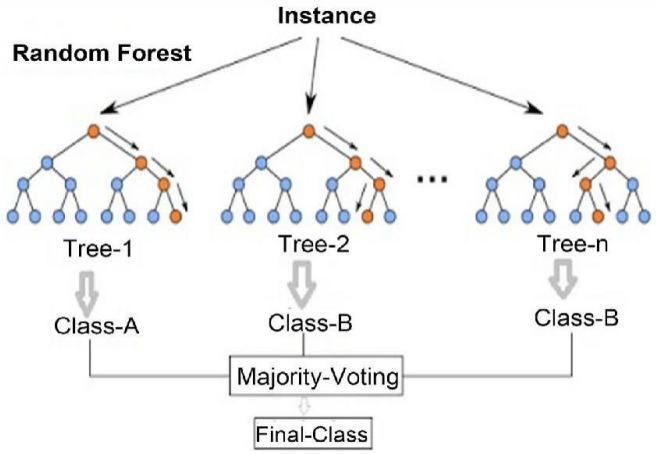

Fig (6): Random Forest processK Nearest Neighbours (KNN)

K Nearest Neighbours (KNN) is a simple yet powerful algorithm used for both classification and regression tasks in machine learning. Imagineyou have a group of friends, and you want to knowif a new person should join your group. KNN works like this: it looks at the characteristics (or features) of the new person, such as age, interests,and hobbies, and then finds the K nearest neighbours from your group who have similarcharacteristics. Neighbours here mean the friends who are most similar to the new person based on those features.

Once KNN has found the nearest neighbours, it makes a decision based on them. For example, if most of the nearest neighbours are friendly and welcoming, KNN might decide that the new person should join your group. The value of K determines how many neighbours are considered when making the decision.

A larger K means considering more neighbours, which can make the decision more stable but might also include more noise from less similar neighbours. Overall, KNN is a simple and intuitive algorithm that can be easily understood, making ita popular choice for beginners in machine learning.

Regression Algorithms:

Regression algorithms are used when the target variable is continuous, meaning it can take on any numeric value within a certain range. Examples of regression algorithms include. Linear Regression, Ridge Regression, Lasso Regression, Support VectorRegression (SVR), Random Forest Regression algorithms are commonly used for tasks such aspredicting house prices, stock prices, and sales forecasts. When it comes to predicting stock prices, we're essentially trying to guess where the prices willgo next. It's a bit like trying to predict the weather, but for the stock market! With machine learning, wehave two handy tools: Support Vector Machine (SVM) and Random Forest classifier.

Linear Regression

Linear regression is like drawing a straight linethrough points on a graph to show the relationship between two variables, like how temperature affects ice cream sales. It helps predict one variable based on the other, using asimple formula (y = mx + b), where 'm' is the slope and 'b' is the intercept.

Ridge Regression

Ridge regression in stock market detection helps manage the risk of overfitting by adding a penalty to the regression coefficients. It shrinks the coefficients, reducing their influence on the model's prediction, which canlead to more stable and reliable results when predicting stock prices or trends. This regularization technique helps prevent extremefluctuations in predictions, making it useful forhandling noisy or high-dimensional data often found in stock market analysis.

Lasso Regression

Lasso regression in stock market detection helps identify the most important features for predicting stock prices by enforcing sparsity inthe model. It encourages some coefficients to be exactly zero, effectively selecting a subset of features while shrinking others. This feature selection capability is valuable for uncovering the most influential factors driving stock market movements, aiding in more interpretable and efficient models.

Support Vector Regression (SVR)

Support Vector Regression (SVR) in stock market detection employs a nonlinear approach to predict stock prices by finding a hyperplane that best fits the data points in a higher-dimensional space. It aims to minimizethe error between actual and predicted stock prices while allowing for a margin of tolerance. SVR is effective in capturing complex patternsand relationships in stock

market data, makingit useful for forecasting price movements and identifying trends with high accuracy.

Random Forest Regression

Random Forest Regression in stock market detection aggregates the predictions of multiple decision trees to provide a robust andaccurate forecasting model. By considering various features and their interactions, it can capture complex patterns in stock market data,leading to improved prediction accuracy. Random Forest Regression is resilient to overfitting and noise, making it well-suited forhandling the volatility and uncertainty inherentin stock market analysis.

PROPOSED SYSTEM

Implementing Support Vector Machine (SVM) algorithm for stock prediction involves several key steps.

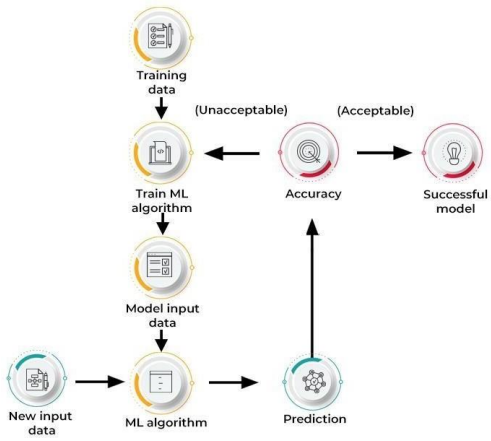

Fig:- PROPOSED SYSTEM

Machine Learning: - Machine learning is a branchof artificial intelligence that focuses on developing algorithms capable of learning from and making predictions or decisions based on data. It involves the construction of models that can automatically discover patterns, associations, and trends in data without being explicitly programmed. Machine learning algorithms are categorized into supervised,unsupervised, and reinforcement learning, each suited to different types of tasks and data. These algorithms find applications across various domains, including healthcare, finance, marketing, and autonomous vehicles, among others. Central tomachine learning is the concept of training models on labelled data to generalize and make predictionson new, unseen data accurately. Feature engineering, data preprocessing, and model evaluation are critical steps in the machine learningpipeline to ensure the robustness and effectiveness of the models. Ethical considerations such as bias, fairness, transparency, and accountability are increasingly important in machine

learning applications, warranting careful attention and mitigation strategies. Additionally, the rapid advancement of machine learning techniques and technologies necessitates ongoing research, education, and collaboration to address emerging challenges and opportunities in this field.

Google Collab

Seamlessly with Jupyter notebooks, allowing users towrite and execute code in a cell-based format with support for markdown, code snippets, and visualizations. Google Collab offers free access to Google Collab, short for Google Collaboratory, is a cloud-based platform provided by Google for writing, executing, and sharing Python code in a collaborative environment. It integrates computational resources such as CPU, GPU, and TPU, making it ideal for running machine learning and data analysis workflows without the need for expensive hardware. It also provides integration with Google Drive for importing and exporting data, along with pre-installed libraries like TensorFlow, PyTorch, NumPy, and pandas, enabling users to work with popular data science tools effortlessly. With features for sharing and collaboration, Google Collab facilitates real-time collaboration on projects,making it a valuable tool for data scientists,researchers, and educators seeking an accessible andpowerful computing environment.

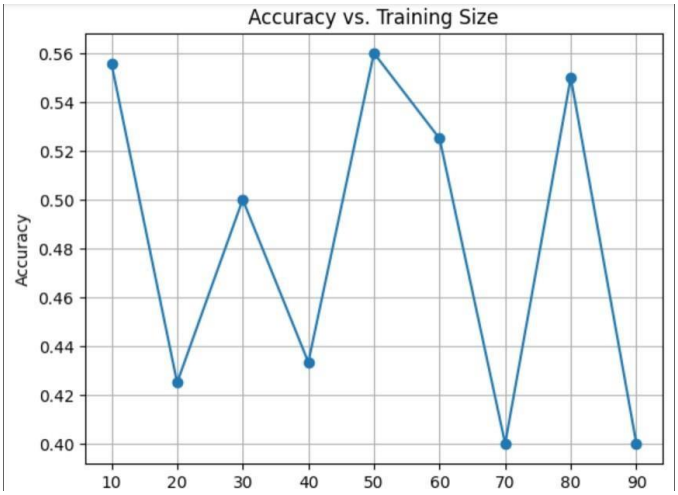

FIG:- Random Forest Predicted Graph

IV. RESULTS

When comparing Support Vector Machines (SVM) and Random Forest for predicting stock prices usingmachine learning, SVM often stands out with its higher accuracy. SVM works by finding the best possible boundary between different classes of data points. In the context of stock price prediction, it triesto find the boundary that best separates the stock price movements into categories like increase, decrease, or stay the same. Random Forest, on the other hand, operates by creating multiple decision

trees and then averaging their predictions. While it'sa powerful technique with its ability to handle large datasets and avoid overfitting, it may not always perform as well as SVM in predicting stock prices accurately.

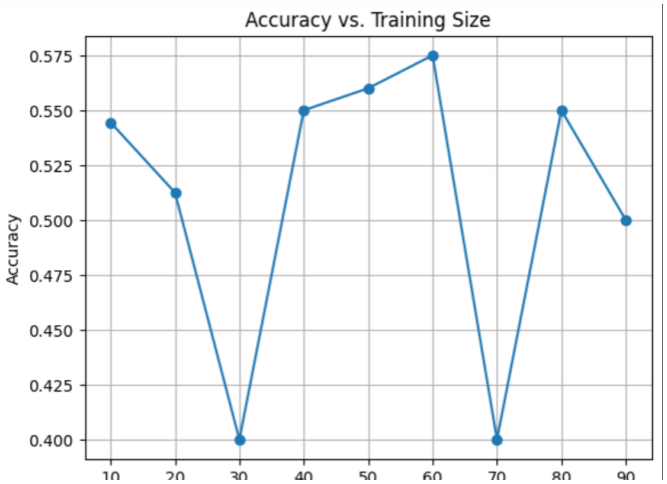

FIG:-SVM Predicted graph

V. CONCLUSIONS

In conclusion, employing the SVM algorithmfor predicting stock prices has showcasedpromising accuracy levels when compared toRandom Forest. SVM's ability to identifyclear boundaries between classes can be advantageous in stock price prediction tasks,particularly when the goal is binary classification, such as predicting whether a stock's price will increase or decrease. The algorithm's effectiveness in handling smaller datasets and its robustness to outliers further contribute to its appeal in this domain.

However, despite SVM's success, thereremains potential for future work to enhancestock price prediction using machine learning techniques.

VI. REFERENCES

1. "Stock price prediction using LSTM,RNN and CNN-sliding window modelIEEE Conference Publication."https://ieeexplore.ieee.org/document/8 126078 (accessed Dec. 27, 2019).
2.]J. Jagwani, M. Gupta, H. Sachdeva, and A. Singhal, "Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend," in 2018 Second International Conference on IntelligentComputing and Control Systems (ICICCS), Madurai, India, Jun. 2018, pp. 462–467, doi: 10.1109/ICCONS.2018.8663035.

3. I. Parmar et al., "Stock Dec. 2018, pp.574–576, doi:10.1109/ICSCCC.2018.8703332.
4. nY. Lei, K. Zhou, and Y. Liu, "Multi-Category Events Driven Stock PriceTrends Prediction," in 2018 5th IEEEInternational Conference on CloudComputing and Intelligence Systems(CCIS), Nanjing, China, Nov. 2018,pp.497–501,doi: 10.1109/CCIS.2018.8691392.
5. B. Jeevan, E. Naresh, B. P. V. Kumar,and P. Kambli, "Share Price Predictionusing Machine Learning Technique," in2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739647.
6. M. Usmani, S. H. Adil, K. Raza, and S.

S. A. Ali, "Stock market prediction using machine learning techniques," in2016 3rd International Conference on computer and Information Sciences (ICCOINS), 2016, pp. 322–327.

7. J. Du, Q. Liu, K. Chen, and J. Wang,"Forecasting stock prices in two waysbased on LSTM neural network," in2019 IEEE 3rd InformationTechnology, Networking, Electronicand Automation Control Conference(ITNEC), Mar. 2019, pp. 1083–1086,doi:10.1109/ITNEC.2019.8729026. Language Processing (EMNLP), Doha,Qatar, Oct. 2014, pp. 1746–1751,doi:10.3115/v1/D14-1181.
8. B. Athiwaratkun and K. Kang, "FeatureRepresentation in ConvolutionalNeural Networks," rXiv150702313 Cs, Jul. 2015, Accessed: Apr. 14,2020.[Online].Available: http://arxiv.org/abs/1507.02313.
9. F. A. Gers, D. Eck, and J. Schmid Huber, "Applying LSTM to TimeSeries Predictable through Time-Window Approaches," in ArtificialNeural Networks — ICANN 2001, Berlin, Heidelberg, 2001, pp. 669–676,Doi: 10.1007/3-540-44668-0_93.
10. J.Han and C. Moraga, "The Influence of the Sigmoid Function Parameters onthe Speed of Backpropagation Learning," in Proceedings of the International Workshop on ArtificialNeural Networks: From Natural to Artificial Neural Computation, Berlin,Heidelberg, Jun. 1995, pp. 195–201, Accessed: Apr. 13, 2020. [Online].
11. "NSE Listed 1000+ Companies' HistoricalData."https://kaggle.com/ab hishekyana/nselisted-1384-companies-data (accessed Jul. 07, 2020).
12. A. Tipirisetty, "Stock Price Predictionusing Deep Learning," p. 60.
13. "Python Numpy Tutorial | Learn Numpy Arrays With Examples," Edureka,Jul. 14, 2017. https://www.edureka.co/blog/python- numpy-tutorial/ (accessed Apr. 21,2020).
14. "Scikitlearn Tutorial: Machine Learning in Python –Dataquest."https://www.dataquest.io/blog/sci-kit-learn-tutorial/ (accessedApr. 21, 2020).
15. J. Brownlee, "Introduction to thePython Deep Learning Library TensorFlow," Machine Learning Mastery,May04,2016.https://machinel earningmastery.com/introductionpytho n-deep-learning-libraryTensorFlow/ (accessed Apr. 21, 2020). [16] "Home – Keras Documentation."https://keras.io/ (accessed Apr. 21, 2020).